Методические и методологические проблемы анализа развития науки

УДК 014.3

Д. В. Золотарёв
канд. экон. наук, заведующий сектором,
Российский научно-исследовательский институт
экономики, политики и права
в научно-технической сфере (РИЭПП),
Москва, Россия, zolotarev@riep.ru

ПОДХОДЫ К РЕШЕНИЮ ПРОБЛЕМЫ ДВОЙНОГО УЧЕТА НАУЧНЫХ СТАТЕЙ ПРИ ОЦЕНКЕ РЕЗУЛЬТАТИВНОСТИ НАУЧНЫХ ОРГАНИЗАЦИЙ

Аннотация

В статье рассмотрены подходы к учету научных статей. Основное внимание уделено методу фракционного счета статей. Определено, что метод пропорционального фракционного учета статей концептуально разделяет научный результат, если статья написана несколькими авторами из разных организаций. Предлагается пропорциональный фракционный метод учета, основанный на нормализации аффилиаций авторов каждой статьи, объединять с равномерным фракционным счетом, основанном на пересечении множеств. Это может дать положительный эффект, дополнив информацию о долях статей, причисляемых организациям, сведениями о наиболее ценных коллаборациях.

Ключевые слова

Методы учета научных статей, пропорциональный фракционный метод учета, равномерный фракционный метод учета, Web of Science, комбинаторика.

D. V. Zolotarev
Candidate of Sciences (PhD) in Economics, Unit Head,
Russian Research Institute of Economics,
Politics and Law in Science and Technology (RIEPL),
Moscow, the Russian Federation, zolotarev@riep.ru

APPROACHES TO SOLVE THE PROBLEM OF DOUBLE-COUNTING SCIENTIFIC ARTICLES WHEN EVALUATING THE EFFECTIVENESS OF SCIENTIFIC ORGANIZATIONS

Abstract

Approaches to the counting of scientific articles are addressed in this paper. Principal attention is paid to fractional counting. It is designated that the method of proportional fractional counting conceptually splits up a scientific result if an article is written by several authors. A combination of proportional fractional counting based on normalization of the affiliations of each article's author(s), and equal fractional counting based on crossing arrays is suggested. This may have a positive effect, combining the information on an organization's share of articles and the most valuable collaborations.

Keywords

Scientific article counting methods, proportional fractional counting, equal fractional counting, Web of Science, combinatorics.

При учете числа научных публикаций, изданных в стране за временной период (например, год), возникает проблема двойного (или большего числа раз) учета одной и той же публикации. Такая ситуация возникает в том случае, если коллектив авторов одной статьи представляет разные научные организации. Для устранения проблемы двойного учета публикаций необходима разработка специальных подходов к общему учету публикаций по научным организациям. Важно отметить, что устранение двойного учета публикаций может существенно повлиять на оценку результативности научных организаций.

Наукометрические методы используются для оценки результативности труда отдельных ученых, для определения степени вовлеченности страны в международное научное сотрудничество, для определения степени научной значимости труда исследователя и т. д.

Проблема измерения результатов научных исследований, в которые вовлечено большое количество ученых (в том числе из разных стран), исследуется довольно давно. Вопрос распределения долей международных научных публикаций между странами рассматривается в работах [1; 2]. Авторами показано, что библиометрическими методами могут быть зафиксированы изменения в международных научных взаимодействиях по разным областям науки в разные периоды времени. Также библиометрические или наукометрические методы активно используются

для оценки труда ученого с целью определения степени эффективности его деятельности и определения уровня оплаты его труда [3].

Несмотря на наличие исследований по учету статей для различных организационных единиц (от отдельного ученого до целых стран), проблема инфляции научных публикаций при подсчете их количества остается не решенной.

Проблема двойного учета статей возникает из-за необходимости организациями сектора исследований и разработок в России заполнять форму отчетности, не соответствующую уровню развития методов учета и оценки результативности научных организаций. Форма утверждена приказом Министерства образования и науки Российской Федерации [4].

Существует несколько способов учета статей [5]. Самым простым считается полный счет, который предполагает, что статья, написанная несколькими авторами, причисляется каждому соавтору целиком. Однако существуют способы фракционного или дробного счета. В этом случае соавторам и, соответственно, организациям, за каждую статью приписывается не по единице публикации, а по ее доле. Если одну статью написали n авторов, то каждому приписывается по 1/n доли статьи.

Современная методология учета статей предполагает наличие трех общих подходов (как для научных организаций, так и для целых стран) [6].

Первый подход – полный счет, когда каждому участнику коллаборации присуждается целая статья. Этот подход объединяет два метода счета: 1) каждому уникальному участнику коллаборации (организация или страна) начисляется один балл; 2) каждому участнику присуждается такое количество баллов, которое соответствует числу авторов от этого участника.

Второй подход – прямой счет. Присуждается один балл только одному автору (это может быть первый автор или автор для корреспонденции, так называемый corresponding author).

Третий подход – фракционный счет. Каждому участнику присуждается доля статьи. Статья может 1) делиться на равные доли в соответствии с количеством уникальных участников коллаборации, или 2) доли могут распределяться в зависимости от количества авторов от каждого участника коллаборации.

В случае расчета числа статей для организаций может применяться как простой долевой счет (в равных долях), так и в долях, пропорциональных числу авторов из соответствующей организации.

Современный инструментарий, которым оснащены базы данных систем научного цитирования, позволяют рассчитывать количество статей для отдельных организаций, авторов, стран и т. д.

Кроме теоретико-методического и методологического интереса способ учета статей представляет интерес с точки зрения органов государственного управления научно-технической сферой. Публикационная активность организаций может повлиять на решение о финансировании, на процедуру оценки грантовых заявок, на возможность привлечения зарубежных партнеров и т. д.

В работе будут рассмотрены два подхода к оценке публикационной активности научных организаций в России, сделаны выводы об их назначении и сути, а также предложены некоторые рекомендации по применению этих подходов в области управления научными исследованиями в России.

Оба подхода основаны на фракционном учете публикаций.

Первый подход описан в работе С. Л. Парфеновой [7]. В ходе устранения множественного учета одной и той же статьи в Российском научно-исследовательском институте экономики, политики и права в научно-технической сфере (РИЭПП) была разработана методика учета статей, предполагающая исключение дублирования их числа.

Основная формула методики позволяет рассчитать число публикаций в долевом (фракционном) выражении, приходящихся на одну статью.

$$N_i = \sum_{j=1}^{n_j} \frac{P_{ij}}{R_j}, \text{ где:} \qquad (1)$$

N_i – число статей организации i, «очищенных» от двойного учета (чистое число статей организации);

n_i – число статей, авторы которых указали организацию i в качестве места работы;

P_{ij} – число связей (аффилиаций) авторов с организацией i, указанных в статье j;

R_j – общее число аффилиаций, указанных всеми авторами статьи j.

Эта формула реализует второй способ учета статей в рамках фракционного подхода, когда доли статьи распределяются в соответствии с количеством авторов от каждого участника коллаборации.

Для иллюстрации работы формулы, объяснения ее сути и возможных последствий ее применения в государственном управлении; расчета научной результативности организаций через публикационную активность, приведем гипотетический пример.

Рассмотрим случай, когда статья написана 5 авторами из 3 организаций. 4 автора указали в качестве места работы по одной организации, и один автор указал 2 организации. Обозначим авторов: «А1», «А2», «А3», «А4», «А5»; организации «Организация1», «Организация2» и «Организация3». Автор «А1» указал «Организацию1», автор «А2» – «Организацию2», автор «А3» – «Организацию3», автор «А4» – «Организацию1», автор А5 – «Организацию1» и «Организацию2».

Занятость автора в той или иной организации есть связь между автором и организацией, называемая аффилиацией. Этим термином будем обозначать здесь отношение между автором и организацией, взятое из метаданных статьи.

Для каждой организации необходимо определить количество аффилиаций (числитель), чтобы узнать долю статьи, приходящуюся на каждую организацию.

Иллюстрация определения аффилиаций для приведенного примера показана на рис. 1.

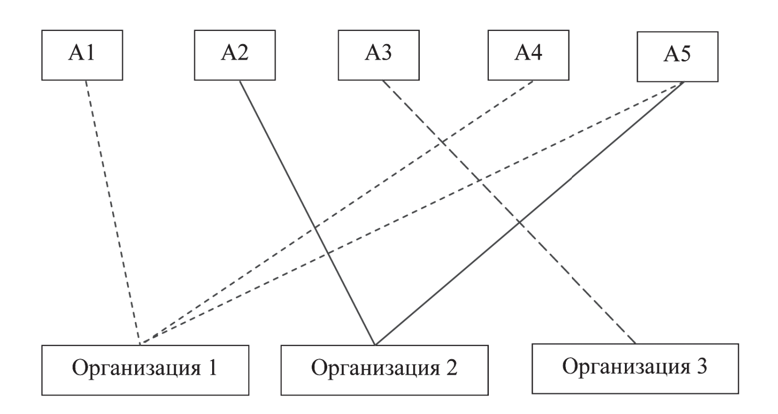


Рис. 1. Интерпретация связей (аффилиаций) авторов и организаций из описания гипотетической статьи

Таким образом видим, что всего аффилиаций в данной статье – 6. На каждую аффилиацию приходится равная доля статьи, в данном случае – 1/6.

Поскольку к «Организации1» приходит 3 аффилиации, то этой организации приписывается 3/6 долей статьи, или половина – 1/2. В «Организацию2» приходят 2 аффилиации, соответственно к ней относится 2/6 долей, или 1/3. На «Организацию3» приходится только одна аффилиация, т. е. 1/6 доля.

В соответствии с формулой 1, рассчет для данного примера по каждой организации будет выглядеть так.

Знаменатель для всех организаций один: $R_j = 6$. Изменяться будет только числитель.

Организация1:

Числитель $P_{ij} = 3$, тогда $N_i = 3/6 = 1/2$.

Организация2:

Числитель $P_{ij} = 3$, тогда $N_i = 2/6 = 1/3$.

Организация3:

Числитель $P_{ij} = 1$, тогда $N_i = 1/6$.

Приведенный пример расчета доли статьи для трех организаций не иллюстрирует полностью работу формулы, поскольку организациями сектора исследований и разработок не может быть написана только одна статья за какой-либо сравнительно длительный период времени (скажем, год). Однако следующие шаги очевидны: рассчитав таким образом долю каждой статьи для каждой организации, следует сложить доли статей каждой организации, чтобы получить общее «чистое» количество статей организации за рассматриваемый период.

Из приведенного примера следует, что для использования формулы необходима специальная организация записи информации о каждой статье, чтобы иметь возможность надежно и верно рассчитать число публикаций для каждой организации.

Что касается последствий ее внедрения в управленческий процесс, то ее широкое применение может привести к следующим последствиям.

Исследователи из разных организаций будут рассматривать коллаборации как невыгодные мероприятия, поскольку логика подхода показывает, что организации приписываются не целые единицы за каждую публикацию, а доли в зависимости от вклада организации в статью.

При получении научных результатов, оформляемых в виде статьи, организации будут стремиться вписать в число авторов статьи по возможности больше сотрудников, чтобы большая доля статьи приписывалась организации.

Последнее может привести к снижению заинтересованности наиболее продуктивных авторов писать статьи в коллаборациях, повышению иждивенческих настроений менее продуктивных авторов; с другой стороны – к усилению тенденции обособления авторов и написания «моноавторских» статей, что может привести к снижению качества работ, поскольку именно в коллаборациях и в совместной работе рождаются наиболее содержательные статьи и ценные идеи.

Еще одно соображение по поводу учета статей с помощью приведенной формулы.

При использовании данной формулы, «очищая» число статей для разных организаций, осуществляется четкое разграничение вклада организации в результативность научной деятельности в целом в стране (или в группе организаций). При этом результаты, полученные в коллаборациях, приписываются долями к отдельным организациям, совместные работы не подчеркиваются, их значимость нивелируется, что происходит за счет механического разрыва совместного вклада разных организаций, неформальных коллективов в получение научного результата. Безусловно, никакого реального «разрыва» или разделения научных результатов не происходит. Однако, следуя заложенной логике справедливости отнесения долей статей организации в соответствии с их вкладом по авторам, разнесение результатов по организациям не способствует усилению коллабораций. Это, безусловно, недостаток подхода.

Похожий подход для учета публикаций может быть реализован через инструментарий, доступный пользователям веб-интерфейса системы научного цитирования Web of Science (далее для удобства – WoS).

Как и в большинстве баз данных, в WoS реализована возможность поиска информации, массивов данных по условиям, которые могут быть комбинированы с помощью логических операторов «И», «ИЛИ», «НЕТ».

Кроме того, в WoS организации при участии специалистов Thomson Reuters могут создать профиль организации (Organizations-Enhanced). Необходимость создавать профили организаций возникает из-за того, что разные авторы из одной и той же организации по-разному приводят названия этой организации (разные аббревиатуры, ошибки в написании или транслитерации и прочее).

В отличие от приведенной формулы, которая оперирует подходом «от частного к общему», описываемый ниже подход основан на принципе «от общего к частному».

Общая задача ставится такой же, как и в случае с приведенной выше формулой: разработать метод для устранения двойного счета статей, опубликованных авторами из организаций сектора исследований и разработок.

Пусть в гипотетической стране **R** функционирует **n** организаций сектора исследований и разработок, сотрудники которых публикуют статьи в научных журналах. Для нахождения «чистого» количества статей каждой организации необходимо нормализовать группы статей, относящихся к каждой организации в зависимости от количества других организаций-соавторов.

В поисковом запросе по организации в WoS можно получить количество статей, которое относится к данной организации. При этом, в перечень статей попадают все статьи, в описании которых есть название данной организации. При формировании запроса по другой организации в перечень статей попадает часть тех, которые удовлетворили условиям первого поискового запроса. Если общее число анализируемых организаций равно трем, то задача сравнительно несложно решается методами математики множеств. Если число организаций больше, чем три, то решение задачи существенно усложняется и становится возможным через построение запросов с использованием логических операторов. Если, например, число организаций равно 7, то для каждой организации необходимо выяснить, сколько совместных статей написали все 7 организаций, сколько совместных статей у каждой организации написано с комбинациями из 6, 5, 4, 3 и двух организаций, а также сколько статей написано сотрудниками организации без соавторства с сотрудниками из других организаций, входящих в эту семерку. Такой анализ требует построения большого количества запросов к WoS.

Для определения общего числа запросов для множества статей из 7 организаций необходимо обратиться к формуле (2) из комбинаторики, которая позволяет определить число всех возможных сочетаний **N** объектов в группы по **k** элементов, причем порядок элементов в группе не существенен.

$$C_N^k = \frac{N!}{k!(N-k)!}, \text{ где} \qquad (2)$$

C_N^k – число комбинаций для группы из k элементов множества, состоящего из N элементов ($N \geq k$);

$N!$ – факториал N ($N! = 1 \times 2 \times 3 \times \ldots \times N$);

В соответствии с данной формулой (2), если $N = 7$, то необходимо рассчитать количество комбинаций для 6 сочетаний: C_7^2, C_7^3, C_7^4, C_7^5, C_7^6, C_7^7. Очевидно, что сочетание C_7^1 стоит несколько обособленно, поскольку количество статей, написанное сотрудниками организации без соавторства с исследователями из других организаций, определяется после

отделения организаций исследуемой группы, предшествуемого поиску числа статей. Если мы ведем речь о поиске в системе WoS, то поисковой запрос осуществляется с помощью логического оператора «НЕТ», который позволяет исключить из поиска статей названия тех организаций, которые могли бы участвовать в написании статьи совместно с той, для которой производится поиск. В то же время, поиск результата для сочетания C_7^1 нужен, чтобы определить общее число запросов в системе WoS, ведь несмотря на то, что «очищение» от дублирования происходит еще до процесса поиска, но формировать запрос и осуществлять поиск все-таки требуется.

Для всех сочетаний общее число запросов, которое необходимо сделать в системе WoS, равно 127 (C_7^2-21, C_7^3-35, C_7^4-35, C_7^5-21, C_7^6-7, C_7^7-1, C_7^1-7).

После построения запросов в системе WoS, возможно определить для каждой организации, сколько статей написано самостоятельно (без соавторства с другими организациями), сколько статей написано в соавторстве с 2-мя, с 3-мя, с 4-мя, с 5-ю и 6-ю организациями. Затем, каждое число делится на количество организаций-соавторов и получается доля статей, которая может быть отнесена к каждой организации.

Очевидны недостатки и этого подхода.

Во-первых, этот подход использует первый метод фракционного счета, что распределяет равные доли участникам коллаборации, без учета их вклада, хотя бы по числу авторов.

Во-вторых, предполагает наличие некоей системы, сформированной базы данных, которая позволяет строить логические запросы. Следовательно, ставит этот метод в зависимость от работы других людей, которая не может быть проконтролирована и при необходимости исправлена.

В-третьих, результаты расчетов по этой системе не совпадут с результатами расчетов по формуле (1).

В-четвертых, если количество организаций не равно 7, а существенно превышает это число (что и есть в реальности), то работа по расчету с помощью этого второго подхода становится слишком трудоемкой.

Такой подход может быть использован для расчета укрупненных групп организаций сектора исследований и разработок.

Например, при реализации политики в научно-технической сфере уполномоченным органам необходимо знать данные о результативности научных исследований на разных уровнях агрегирования. Для быстрого и надежного расчета с использованием системы WoS, могут быть созданы профили организаций для органов исполнительной власти, подведомственные организации которых осуществляют исследования и разработки отдельных крупных вузов и научных институтов, а также групп вузов и научных институтов.

Несмотря на то, что второй подход существенно уступает первому в справедливости отнесения долей статей к той или иной организации, он имеет возможность устранить существенный недостаток первого подхода, а именно разделение научного результата, его концептуальный

разрыв и отнесение совместно полученного цельного научного результата долями по разным организациям.

При осуществлении поисковых запросов в рамках второго подхода, когда анализируется совокупность интересующих организаций, появляется возможность увидеть группы, неформальные объединения, налаженные связи между организациями сектора исследований и разработок. Если к такому анализу добавить дисциплинарные разрезы, то можно увидеть не только то, в каких науках происходят объединения научных коллективов, но и тенденции междисциплинарных исследований.

Каждый из существующих методов учета статей на уровне организаций и стран характеризуется недостатками. Фракционный счет является наиболее справедливым и конструктивным с точки зрения учета статей для оценки результативности научных организаций.

В то же время концептуальная недостаточность каждого из двух методов фракционного счета позволяет говорить о необходимости их развития по траектории сближения.

Для более правильного учета статей с точки зрения повышения эффективности и отдачи от работы организаций сектора исследований и разработок, необходимо начинать учет с разделения научного результата и определения вклада каждой организации отдельно. Но на этом нельзя останавливаться и воспринимать полученную информацию как руководство к действию. Следующим шагом должно стать определение степени и эффективности совместной работы этих организаций. Необходимо на основе полученной информации определять, как усиливают друг друга эти организации, как они взаимодействуют друг с другом.

Будущие исследования

Проведенное исследование позволяет говорить о том, что анализ публикационной активности позволяет получать необходимую информацию для оценки научной деятельности организаций. В то же время, существуют серьезные недостатки в степени оценки совместной работы организаций сектора исследований и разработок. Требуется разработка методов учета совместной деятельности организаций и способов выявления наиболее перспективных коллабораций, в рамках которых могут появиться опережающие исследования. Для возможности получения таких данных необходимо объединять подходы расчета количества публикаций с анализом цитируемости и контекстным анализом публикаций.

Литература

1. *De Lange C., Glänzel W.* Modelling and measuring multilateral co-authorship in international scientific collaboration. Pt. I: Development of a new model using a series expansion approach // Scientometrics. 1997. № 40 (3). P. 593–604.

2. *Egghe L.* An explanation of the relation between the fraction of multinational publications and the fractional score of a country // Scientometrics. 1999. № 45 (2). P. 291–310.
3. *Глаголев М. В., Лапшина Е. Д.* Методика расчета эффективности научной деятельности в… // ДОСиГИК. 2012. Т. 3. № 2 (6).
4. Приказ Министерства образования и науки Российской Федерации от 05.03.2014 № 162 «Об утверждении порядка предоставления научными организациями, выполняющими научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, сведений о результатах деятельности и порядка подтверждения указанных сведений федеральными органами исполнительной власти в целях мониторинга, порядка предоставления научными организациями, выполняющими научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, сведений о результатах деятельности в целях оценки, а также состава сведений о результатах деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, предоставляемых в целях мониторинга и оценки» // Официальный сайт Российской газеты. URL: http://www.rg.ru/2014/05/14/minobrnauki2-dok.html (дата обращения: 29.09.2015).
5. *Писляков В. В*. Библиометрические индикаторы в ресурсах Thomson Reuters. Глава в книге «Руководство по наукометрии: индикаторы развития науки и технологии; монография // *Акоев М. А., Маркусова В. А., Москалева О. В., Писляков В. В.* Руководство по наукометрии: индикаторы развития науки и технологии / Под. ред. М. А. Акоева. Екатеринбург: Изд-во Урал. ун-та, 2014. 250 с. С. 80.
6. *Huang M.-H., Lin C.-S., Chen D.-Z.* Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact // Journal of the American Society for Information Science and Technology. 2011. № 62 (12). P. 2427–2436.
7. *Парфенова С. Л.* Анализ и методика учета публикаций российских исследователей в системе государственного управления на основе данных информационно-аналитических сервисов Web of Science // 4-я Междунар. науч.-практ. конф. «Научное издание международного уровня – 2015: Современные тенденции в мировой практике редактирования, издания и оценки научных публикаций» / Под ред. О. В. Кирилловой. 2015. СПб.: Сев.-Зап. ин-т упр. – фил. РАНХиГС. С. 8.